Bayesian inversion and uncertainty estimation: implications for simulation codes

Kenneth M. Hanson

Los Alamos National Laboratory

Taken from presentations available under http://home.lanl.gov/kmh/

Overview

- Bayesian tomographic reconstruction from two views
  - deformable geometric models with smoothness prior
  - uncertainty characterized by posterior probability distribution
- Markov Chain Monte Carlo (MCMC) technique
  - for drawing random samples from probability density functions
  - tool for estimating and visualizing uncertainties in models
- Optical tomography
  - inversion of time-dependent diffusion process
  - adjoint differentiation of solution to PDEs
- Uncertainties in simulation predictions

Bayesian approach to model-based analysis

- **Models**
  - ▸ used to describe and analyze physical world
  - ▸ parameters inferred from data
- **Bayesian analysis**
  - ▸ uncertainties in parameters described by probability density functions (pdf)
  - ▸ prior knowledge about situation may be incorporated
  - ▸ quantitatively and logically consistent methodology for making inferences about models
  - ▸ open-ended approach
    - can incorporate new data
    - can extend models and choose between alternatives

Bayesian viewpoint

- Focus on probability distribution functions (pdf)
  - uncertainties in estimates more important than the estimates themselves
- Bayes law: $p(\mathbf{a}|\mathbf{d}) \sim p(\mathbf{a})\, p(\mathbf{d}|\mathbf{a})$
  - where **a** is parameter vector and **d** represents data
  - pdf before experiment, $p(\mathbf{a})$ (called *prior*)
  - modified by pdf describing experiments, $p(\mathbf{d}|\mathbf{a})$ (*likelihood*)
  - yields pdf summarizing what is known, $p(\mathbf{a}|\mathbf{d})$ (*posterior*)
- Experiment should provide decisive information
  - posterior distribution much narrower than prior

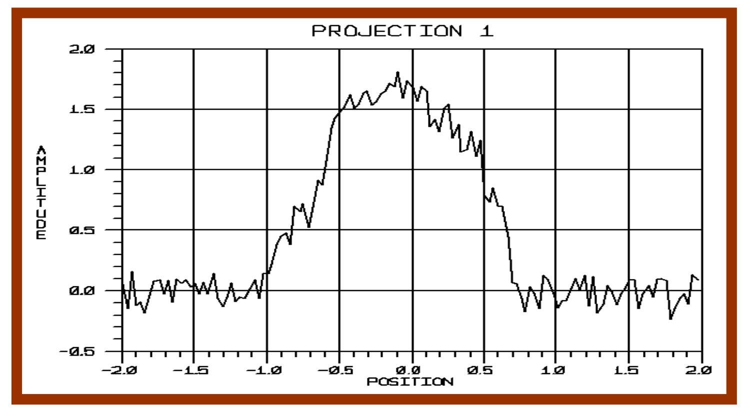
Bayesian model building

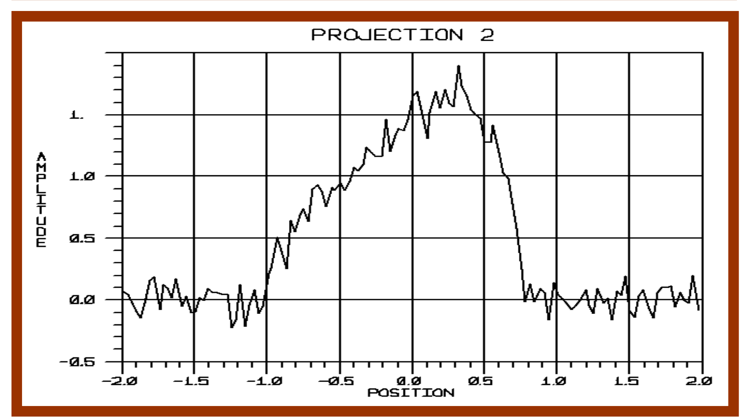
- Steps in model building
  - choose how to model (represent) object
  - assign priors to parameters based on what is known beforehand
  - for given measurements, determine model with highest posterior probability (MAP)
  - assess uncertainties in model parameters
- Higher levels of inference
  - assess suitability of model to explain data
  - if necessary, try alternative models and decide among them

Example - tomographic reconstruction

- Problem - reconstruct object from two projections
  - ▸ 2 orthogonal, parallel projections (128 samples in each view)
  - ▸ additive Gaussian noise with rms dev. = 5% of proj. max

Original object





Likelihood

- Likelihood defined as $p(\mathbf{d}|\mathbf{a})$ = probability of data **d**, given model and its parameters **a**
- For measurements subject to additive, independent Gaussian-distributed noise, minus-log-likelihood is

$$-\log[\,p(\mathbf{d}|\mathbf{a})] = \varphi(\mathbf{a}) = \tfrac{1}{2}\chi^2 = \tfrac{1}{2}\sum \frac{(d_i - d_i^*)^2}{\sigma^2}$$

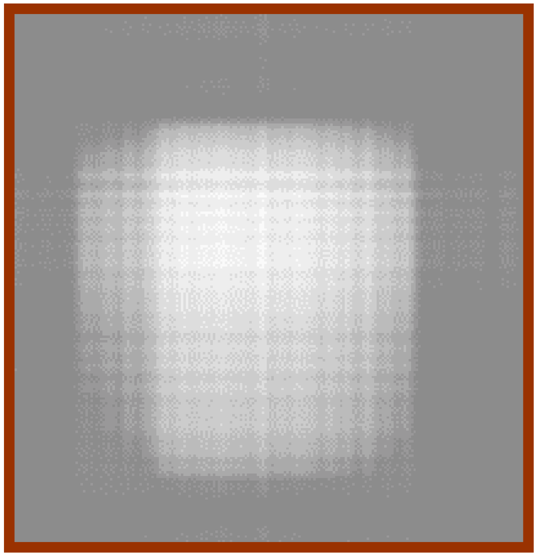
  - where d_i is the ith measurement,
    d_i^* is its predicted value (for specific **a**),
    σ is rms noise in measurements

Standard reconstruction approaches

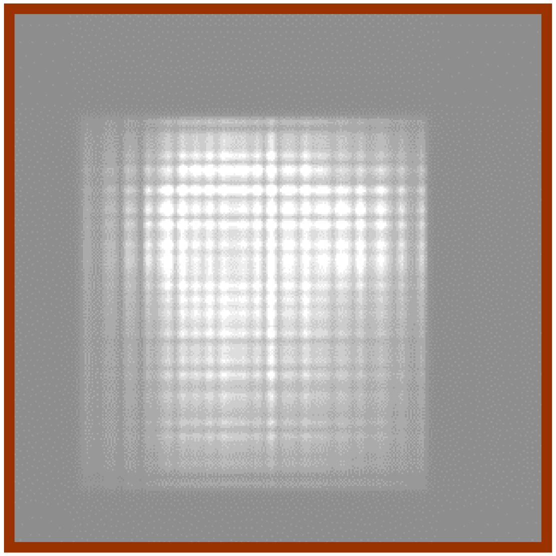
- “Standard” reconstruction algorithms
  - based on minimizing minus-log-likelihood ($\frac{1}{2}\chi^2$) using additive or multiplicative updates, non-negativity constraint
  - do not reproduce original image

Original object

Additive-update reconstruction



Maximum-entropy reconstruction

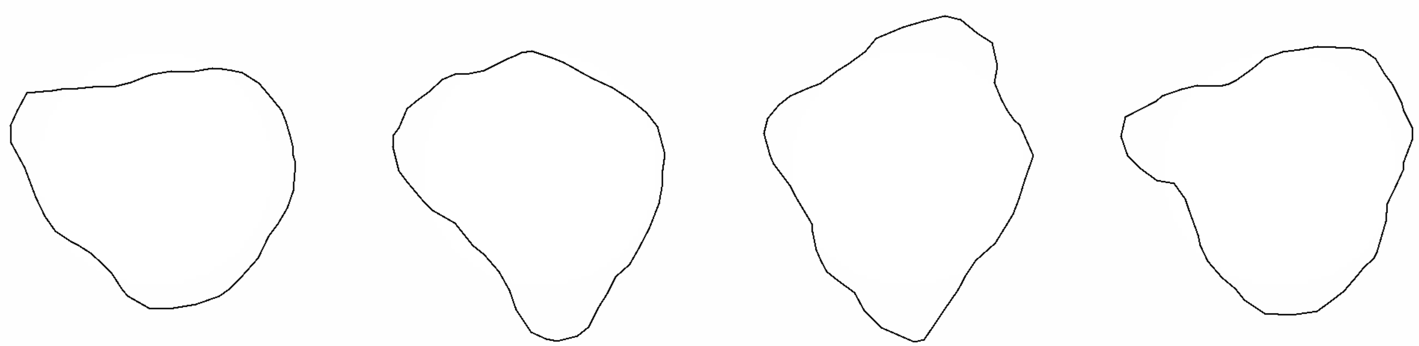


Model-based Bayesian reconstruction - make use of prior information

- Assumptions about object
  - interior density is uniform
  - abrupt change in density at boundary
  - boundary is relatively smooth
- Object model chosen to incorporate these assumptions
  - object boundary - deformable geometric model
  - boundary smoothness achieved through prior
  - interior has uniform density (known)
  - exterior density is zero
  - only variables are those describing boundary

Probabilistic interpretation of prior for deformable boundary model

- Probability of shape: $\sim \exp\left[-\frac{\alpha S}{(2\pi)^2} \oint \kappa^2 ds \right]$
  - where κ = boundary curvature
- Sample prior pdf using MCMC
  - shows variety of shapes deemed admissible before experiment, capturing our uncertainty about shape
  - decide on α = 5 on basis of appearance of shapes



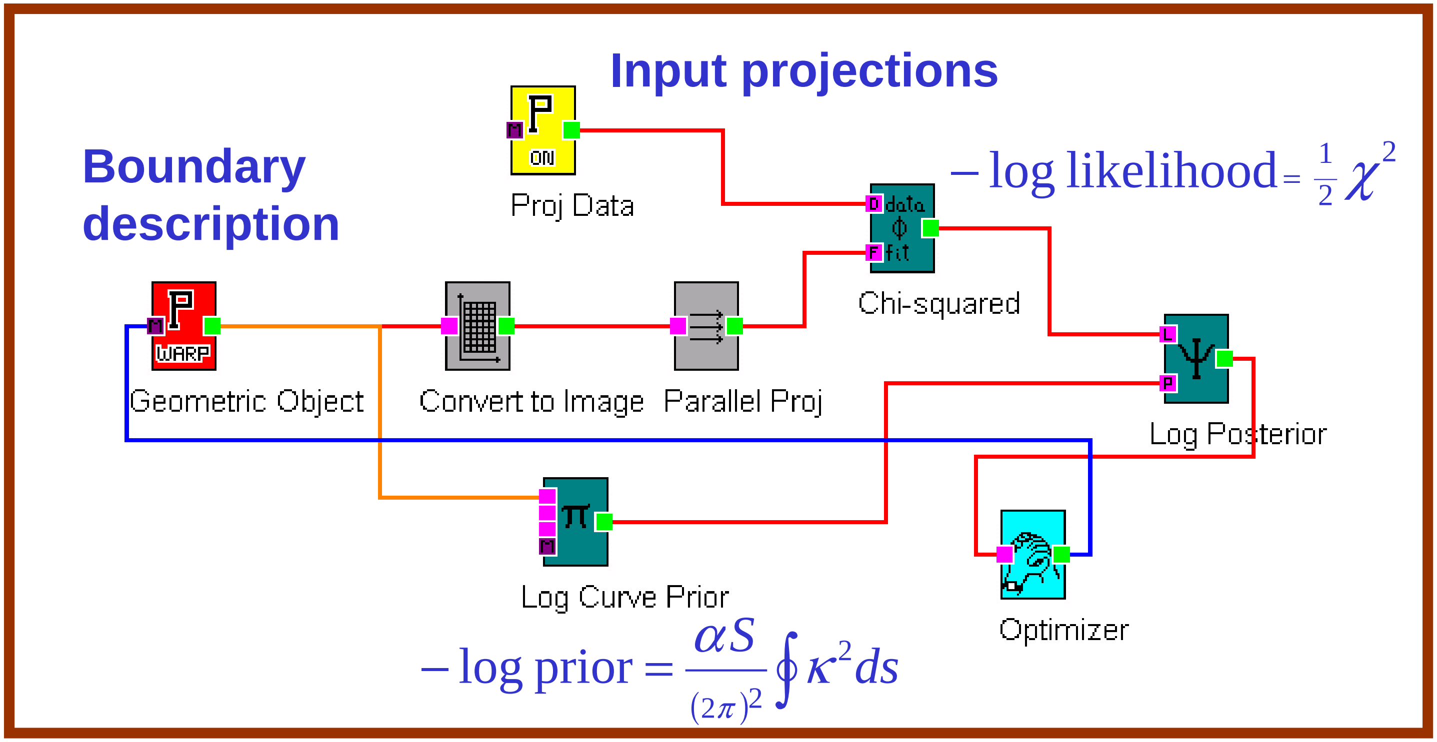
Plausible shapes drawn from prior for α = 5

The Bayes Inference Engine

- Flexible modeling tool developed in DX-3
  - object described as composite geometric and density model
  - measurement process (principally radiography)
- User interface via graphically-programmed data-flow diagram
- Full interactivity through Object-Oriented design
- BIE provides
  - MAP estimate by optimization
  - gradient calculated by adjoint differentiation
  - random samples of posterior by MCMC
  - uncertainty estimates

The Bayes Inference Engine

- BIE data-flow diagram to find MAP solution



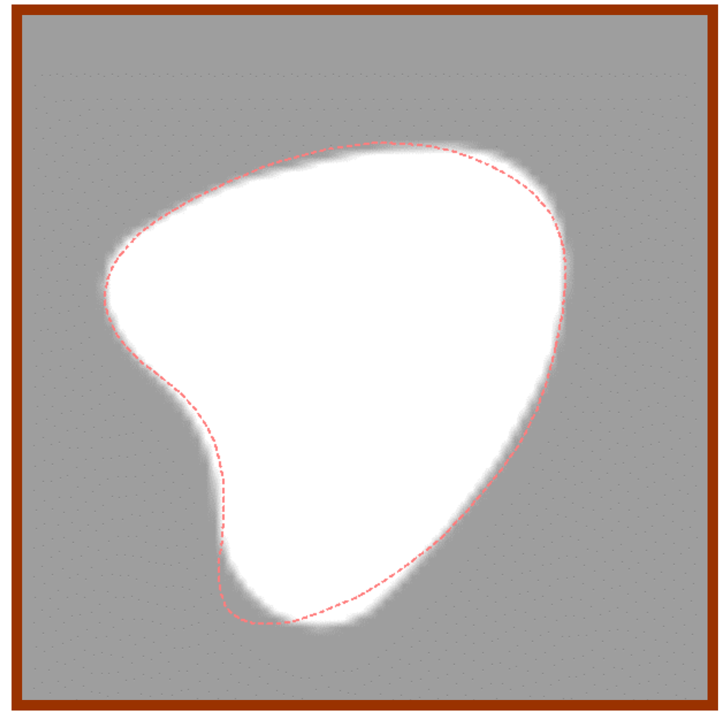
- Optimizer uses gradients that are efficiently calculated by adjoint differentiation in code technique(ADICT)

MAP reconstruction

- Determine boundary that maximizes posterior probability (for $\alpha = 5$)

- Result not perfect, but very good for only two projections

- Question: "How do we quantify uncertainty in reconstruction?"

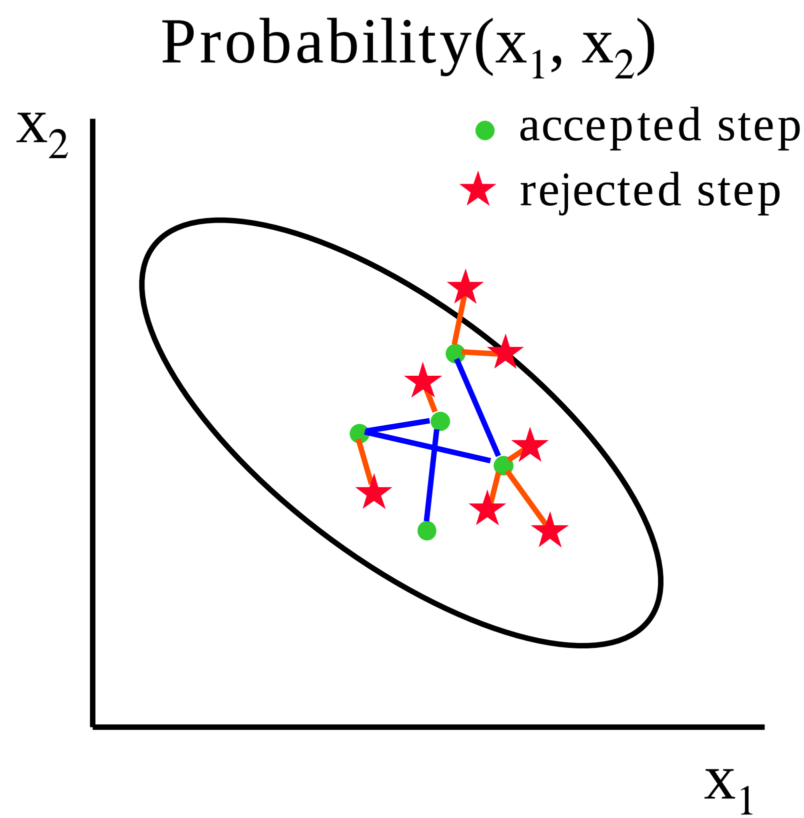


Reconstructed boundary (gray-scale) compared with shape of original object (red line)

Markov Chain Monte Carlo

Generates sequence of random samples from an arbitrary **computed** probability density function

- Metropolis algorithm:
  - draw trial step from symmetric pdf, i.e., $t(\Delta \mathbf{x}) = t(-\Delta \mathbf{x})$
  - accept or reject trial step on basis of probability at new position rel. to old
  - simple and generally applicable
  - relies only on **calculation** of target pdf for any **x**



Uses of MCMC

- Permits evaluation of expectation values of $q(\mathbf{x})$
  - for K samples, $\langle f(\mathbf{x})\rangle = \int f(\mathbf{x})\, q(\mathbf{x})\, d\mathbf{x} \cong (1/K)\, \Sigma_k\, f(\mathbf{x}_k)$
  - typically used to calculate mean $\langle \mathbf{x} \rangle$ and variance $\langle (\mathbf{x} - \langle \mathbf{x} \rangle)^2 \rangle$
- Useful for evaluating integrals, such as the partition function for properly normalizing the target pdf
- Dynamic display of sequence as video loop
  - provides visualization of uncertainties in model and range of model variations
- Automatic marginalization
  - when considering any subset of parameters of an MCMC sequence, the remaining parameters are marginalized over

MCMC Issues

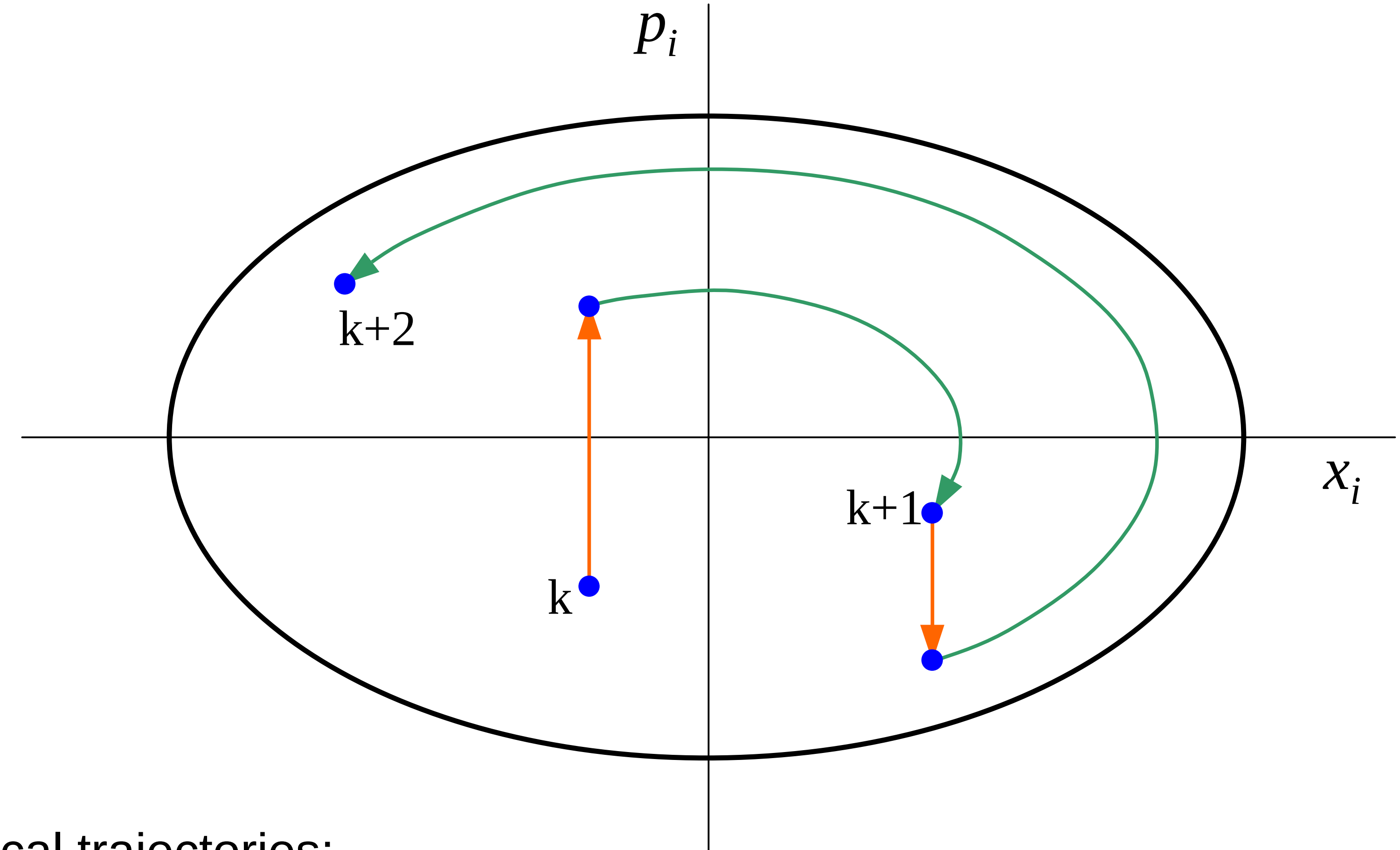
- Confirmation of **convergence** to target pdf
  - is sequence in thermodynamic equilibrium with target pdf?
  - validity of estimated properties of parameters (covariance)
- **Burn in**
  - at beginning of sequence, may need to run MCMC for awhile to achieve convergence to target pdf
- Use of **multiple sequences**
  - different starting values can help confirm convergence
  - natural choice when using computers with multiple CPUs
- **Accuracy** of estimated properties of parameters
  - related to efficiency, described above
- Optimization of **efficiency** of MCMC

Hamiltonian hybrid algorithm

- called hybrid because it alternates Gibbs & Metropolis steps
- associate with each parameter x_i a fictitious **momentum** p_i
- define a Hamiltonian
  $H = \varphi(\mathbf{x}) + \Sigma\, p_i^2/(2\, m_i)$; $\varphi = -\log(q(\mathbf{x}))$; $q(\mathbf{x})$ = target distr.
- construct a new pdf:
  $q'(\mathbf{x}, \mathbf{p}) = \exp(-\, H(\mathbf{x}, \mathbf{p})) = q(\mathbf{x}) \exp(-\Sigma\, p_i^2/(2\, m_i))$
- can easily move long distances in ($\mathbf{x}$, $\mathbf{p}$) space at **constant *H*** using **Hamiltonian dynamics**; so Metropolis step is very efficient
- requires gradient* of φ (minus-log-prob)
- Gibbs step: draw **p** from known Gaussian pdf (at fixed **x**)
- efficiency may be better than Metropolis for large dimensions

* **adjoint differentiation** provides efficient gradient calculation

Hamiltonian hybrid algorithm



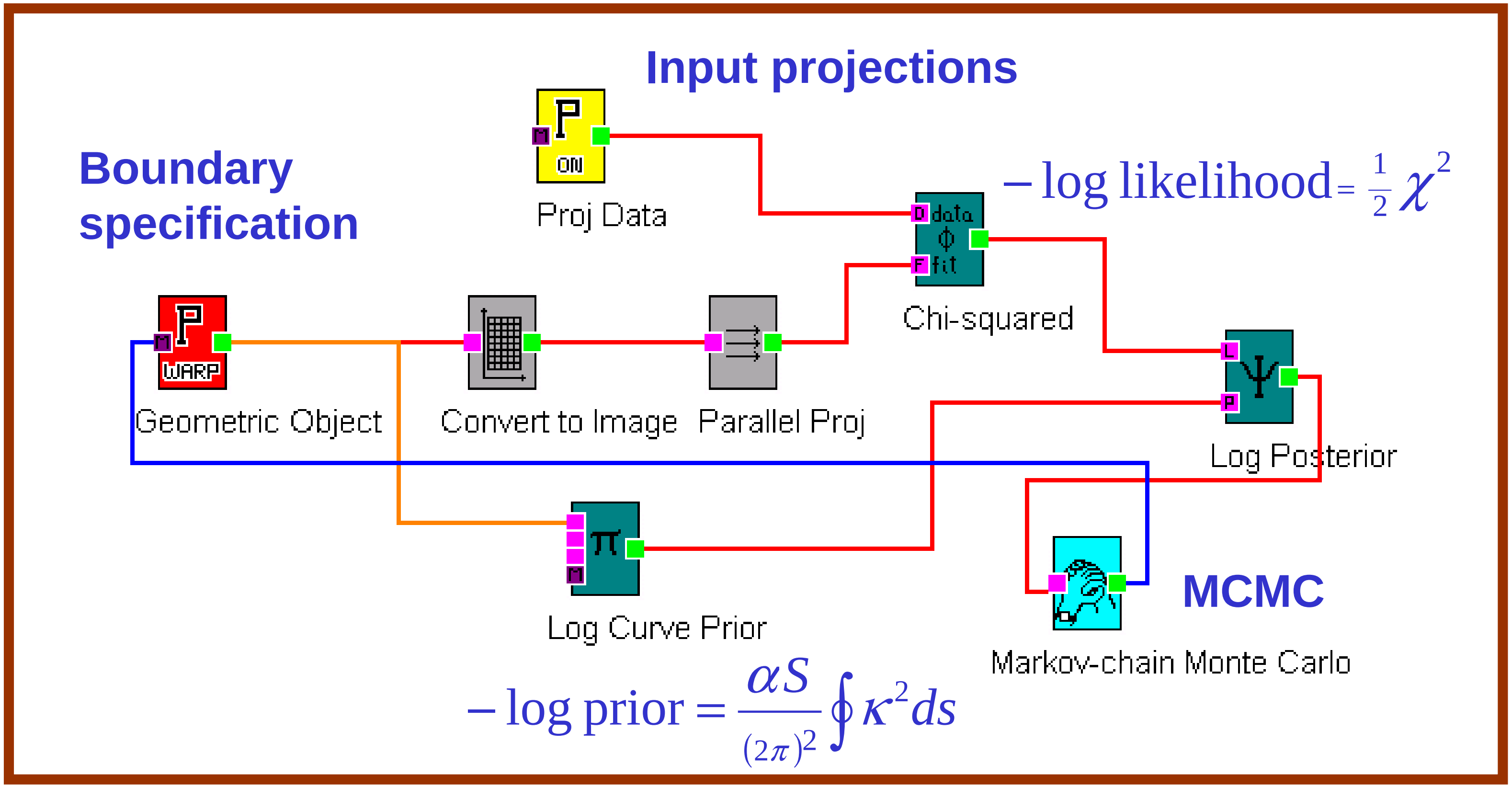
Typical trajectories:

red path - Gibbs sample from momentum distribution

green path - trajectory with constant *H*, followed by Metropolis

The Bayes Inference Engine

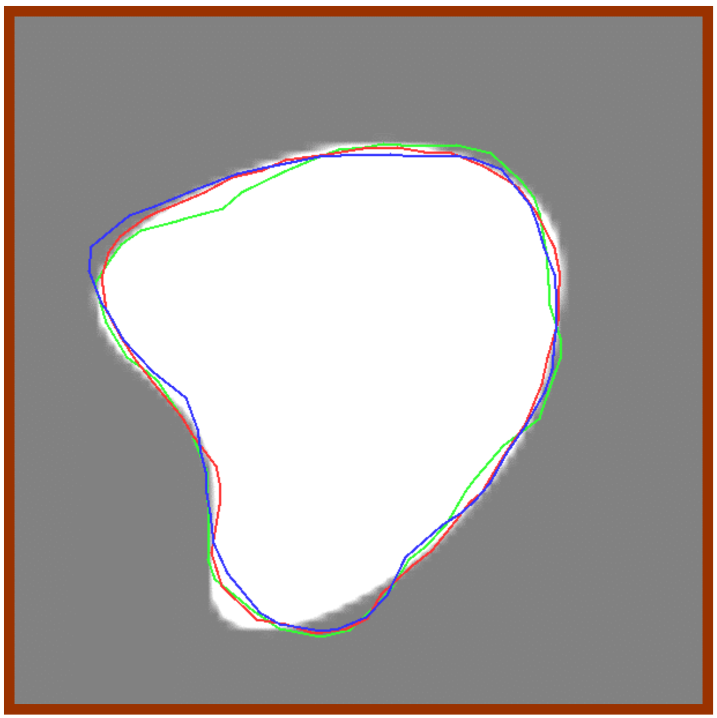
- BIE data-flow diagram to produce MCMC sequence



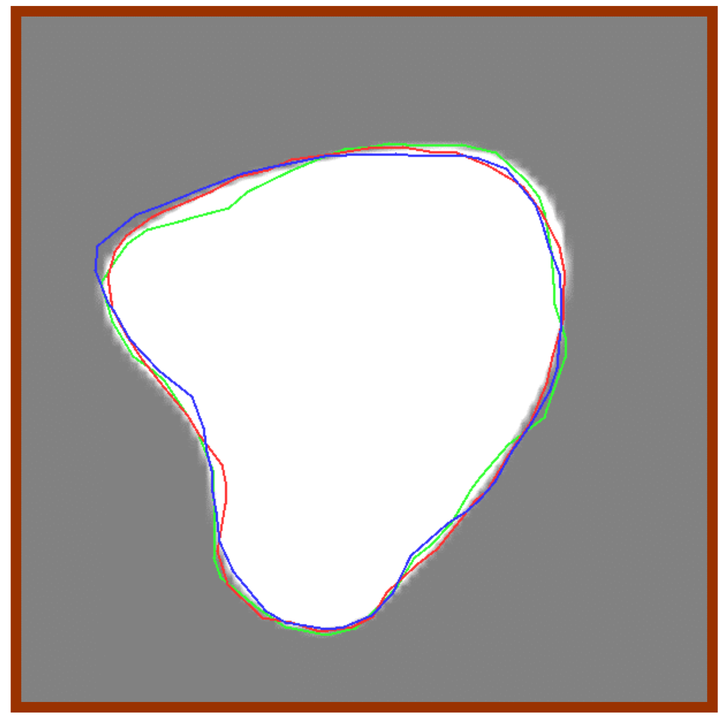
- MCMC module implements Metropolis algorithm

Uncertainties in two-view reconstruction

- From MCMC samples from posterior with 150,000 steps, display three selected boundaries
- These represent alternative **plausible solutions**



compared to original object

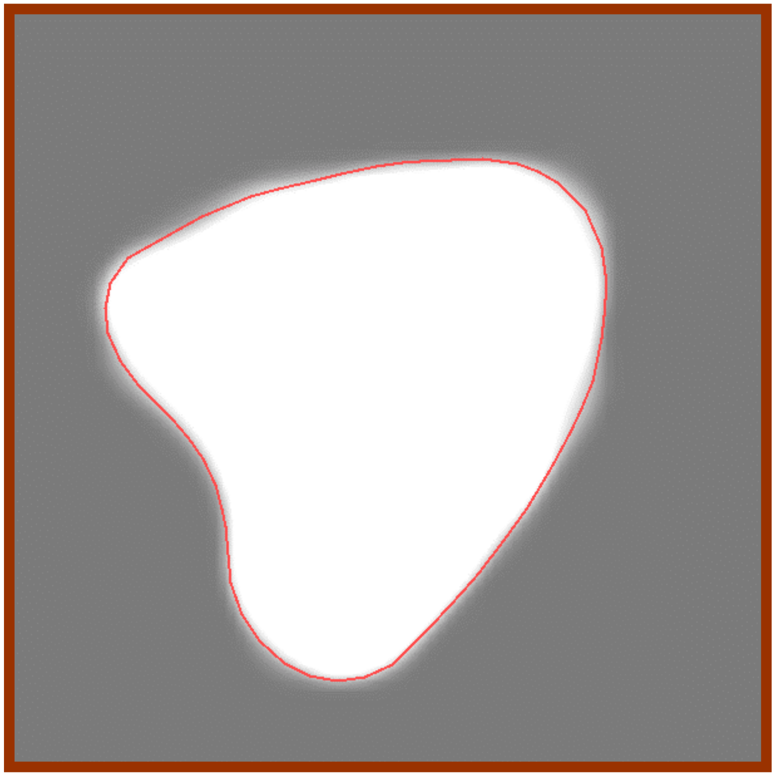


compared to MAP estimated object

Posterior mean of gray-scale image

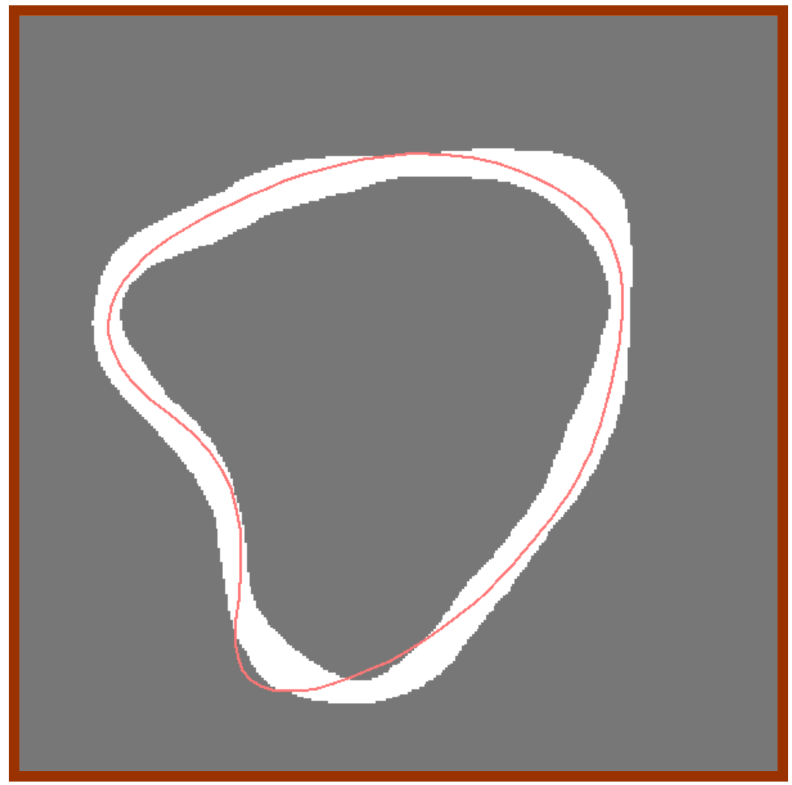
- Average gray-scale images over MCMC samples from posterior
- Value of pixel is probability it lies inside object boundary
- Amount of blur in edge is related to magnitude of uncertainty in edge localization

Posterior mean image
compared to
MAP boundary (red line)



Credible interval

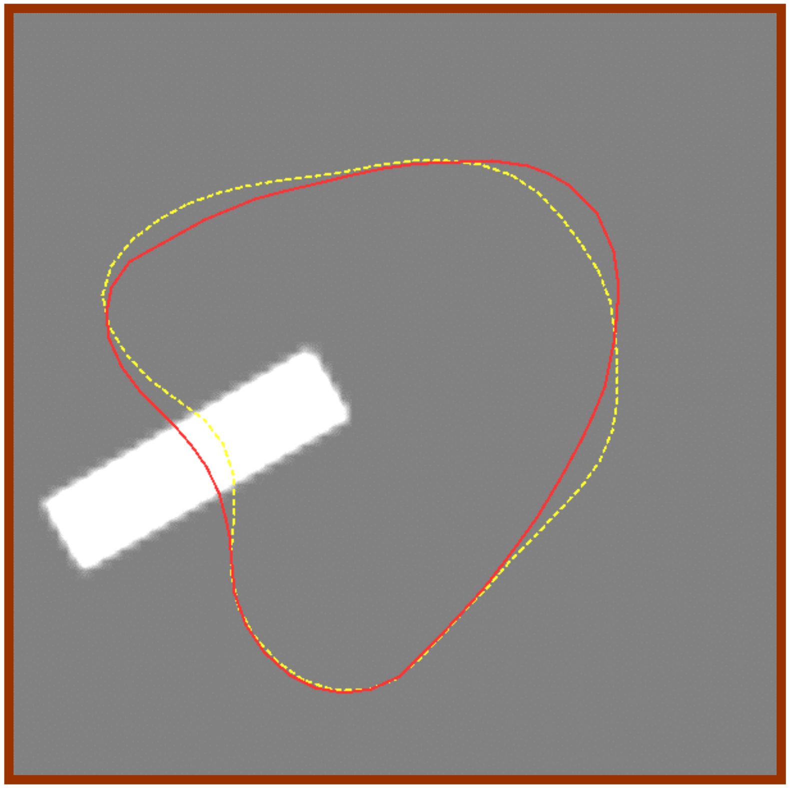
- 95% credible interval of boundary localization for two-view reconstruction compared with original object boundary (red line)
  - narrower at tangent points
  - 92% of original boundary lies inside
    95% credible interval
- Marginalized measure of uncertainty - **ignores correlations** among different positions



Stiffness of posterior related to uncertainty

- Interpret φ = - log probability as potential function; sum of
  - deformation energy (prior)
  - $\frac{1}{2}\chi^2$ (likelihood)
- Stiffness of model proportional to curvature of φ
- Displacement obtained by applying a force to MAP model and reminimizing φ proportional to force times **covariance matrix** (for Gaussian approximation)

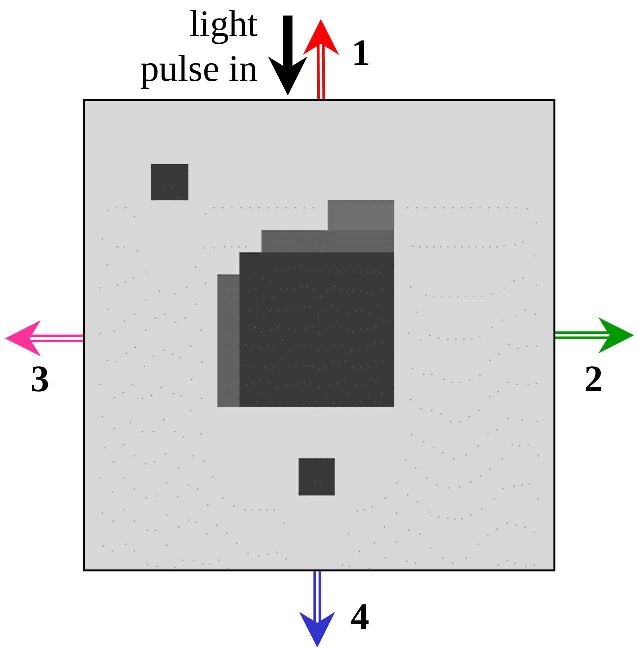
Applying force (white bar) to MAP boundary (red) moves it to new location (yellow-dashed)

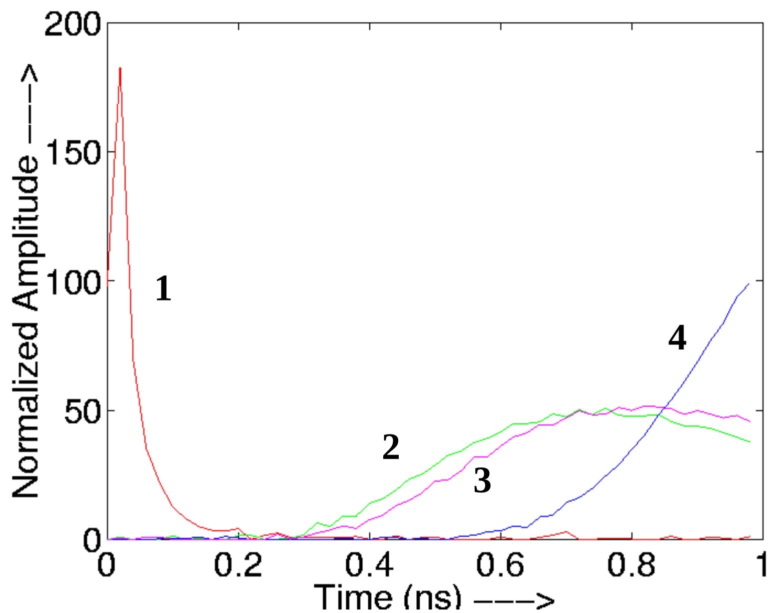


Inversion of complex simulations

- Advanced techniques are required to cope with large data structures and models with numerous parameters
  - Optimization
    - gradient-based quasi-Newton methods (e.g., CG, BFGS)
    - adjoint differentiation for efficient calculation of gradients
    - multiscale methods for controlling optimization process
  - Bayesian methods
    - overcome ill posedness of inversion through use of prior knowledge
    - Markov chain Monte Carlo to characterize uncertainties
  - Appropriate higher-order models
    - Markov random fields
    - deformable geometrical models
    - but also consider lowest order, elemental representations

Simulation of light diffusion in tissue



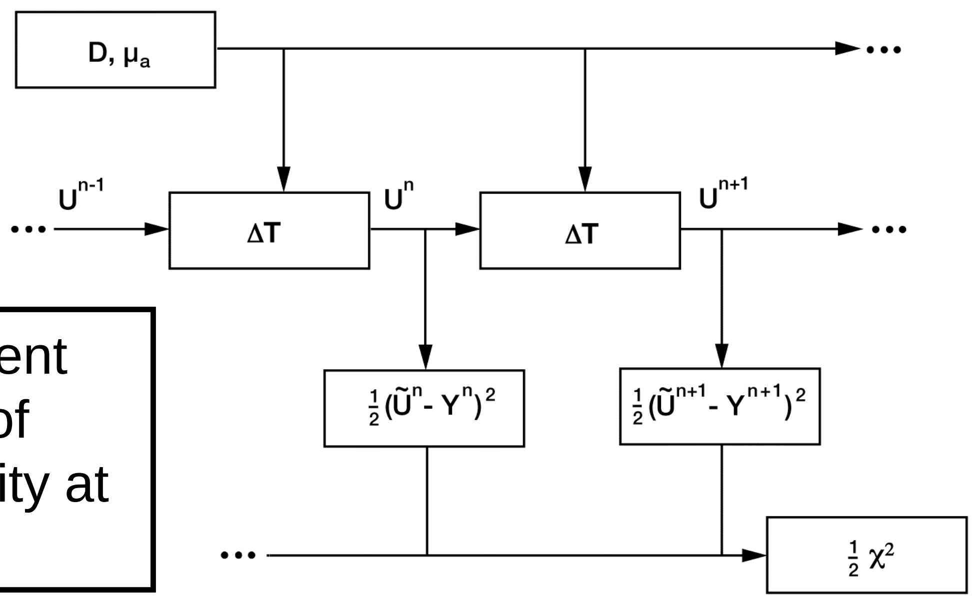


0.7< D <1.4 cm^2ns^{-1} (μ_a=0.1 cm^{-1})

- for assumed distribution of diffusion coefficients (left)
- predict time-dependent output at four locations (right)
- **reconstruction problem** - determine image on left from data on right

Time-dependent finite-difference calculation

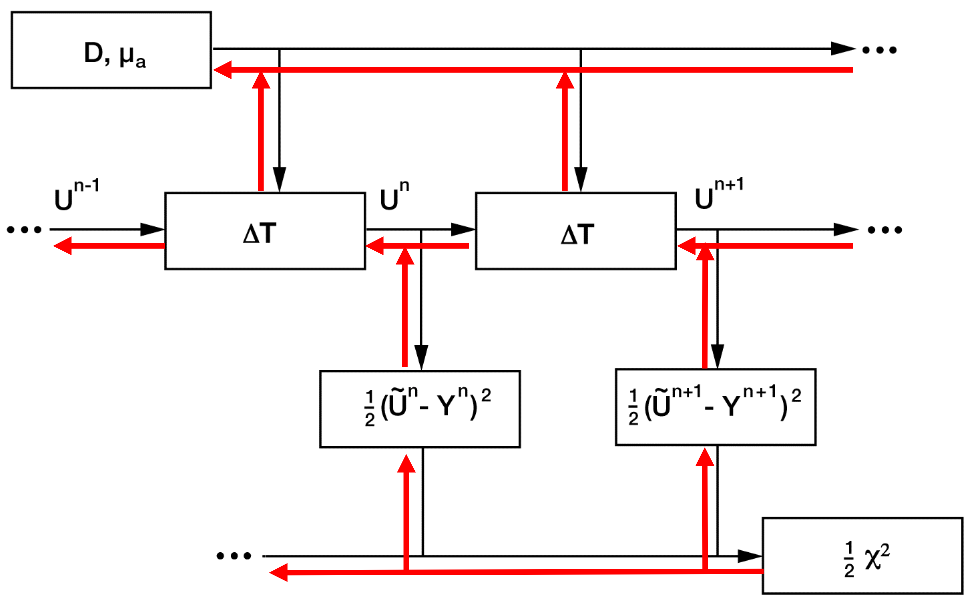
- Data-flow diagram shows calculation of time-dependent measurements by finite-difference simulation
- Calculation marches through time steps Δt
  - new state $\mathbf{U}_{n+1}$ depends only on previous state $\mathbf{U}_n$



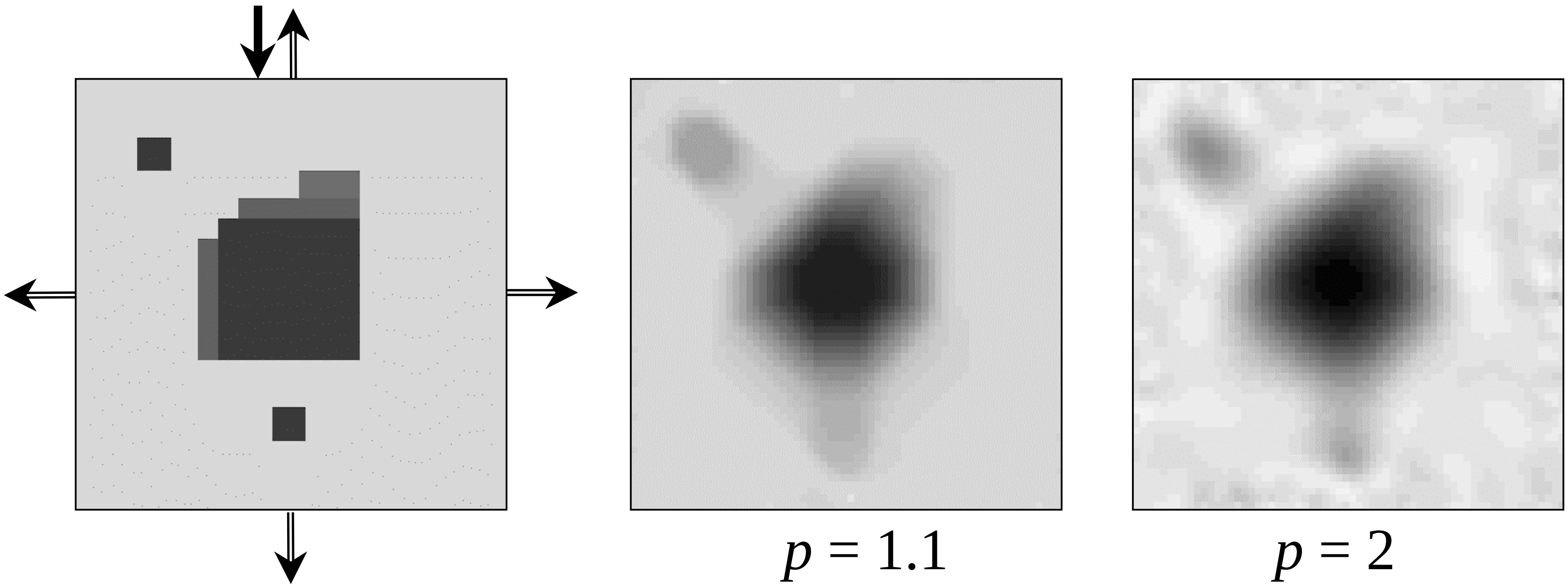
Time-dependent calculation of U_n= light intensity at time n

Adjoint differentiation of forward calculation

- Adjoint differentiation calculation precisely reverses direction of forward calculation
- Each forward data structure has an associated derivative
  - $\mathbf{U}_n$ propagates forward, $\dfrac{\partial \varphi}{\partial \mathbf{U}_n}$ goes backward $(\varphi = \frac{1}{2}\chi^2)$



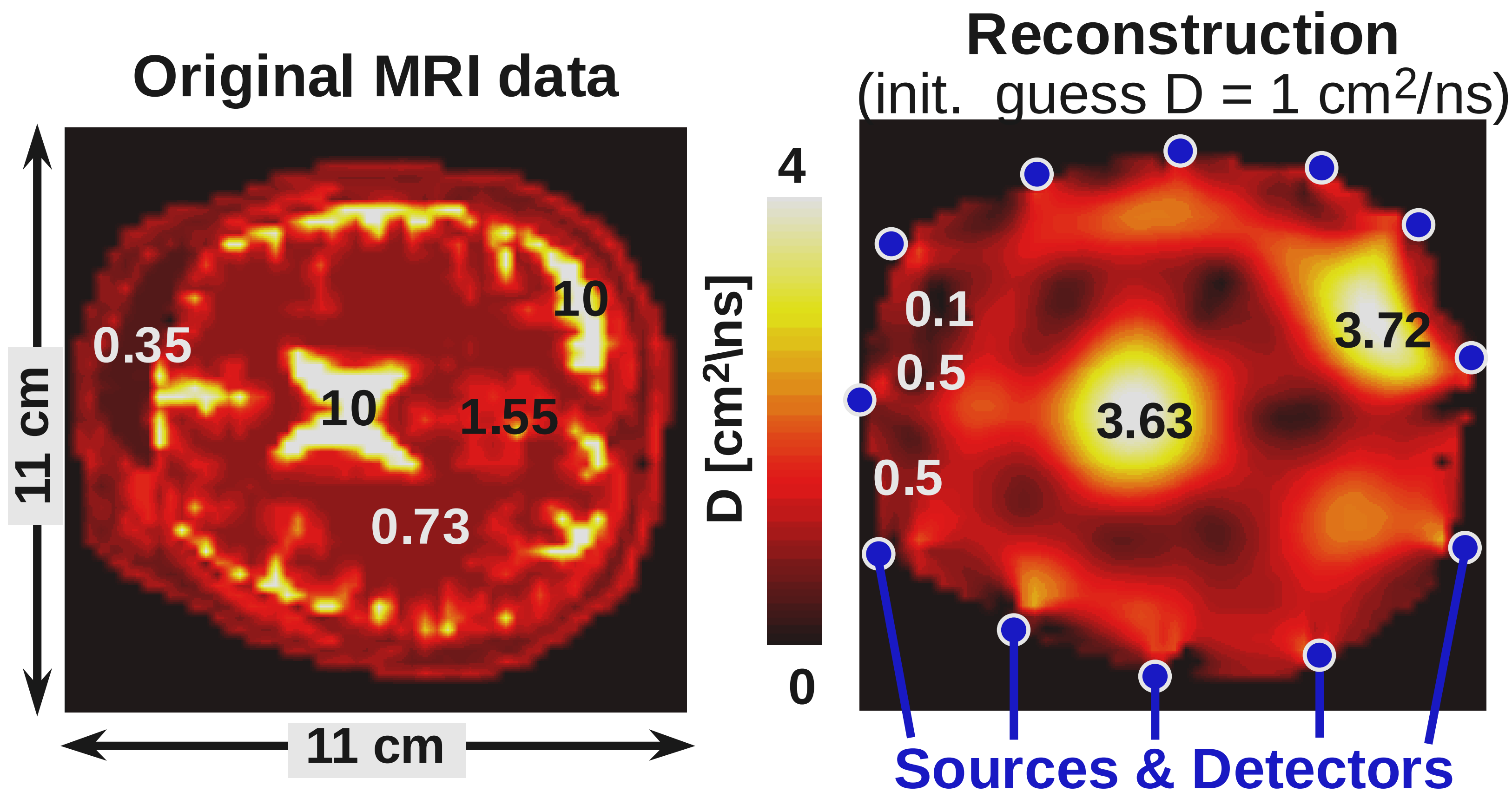
Reconstruction of simple phantom



$p = 1.1$ $p = 2$

- Measurements
  - section is $(6.4\text{cm})^2$, $0.7 < D < 1.4$ $\text{cm}^2\text{ns}^{-1}$ ($\mu_{abs} = 0.1$ cm^{-1})
  - 4 input pulse locations (middle of each side)
  - 4 detector locations; intensity measured every 50 ps for 1 ns
- Reconstructions on 64 x 64 grid from noisy data (rmsn = 3%)
- Prior based on Markov random field with adjustable Lp norm

Reconstruction of Infant's Brain I



hematoma (left side) and cerebrospinal fluid pocket (upper right)

(60 iterations ~ 70 min)

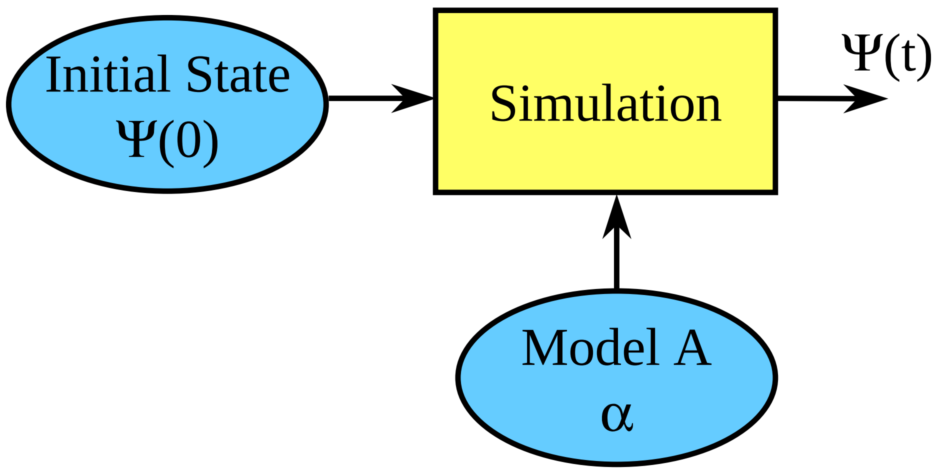
Applications of adjoint differentiation

- Imaging through refractive, reflective, diffusive media
  - seismology, medical and NDE ultrasound, ...
- Sensitivities in large-scale simulations (data assimilation):
  - atmosphere models (Ron Errico, NCAR; Bob Fovell, UCLA)
  - fluid dynamics; hydrodynamics (Rudy Henninger)
- Optimization in large engineering design problems:
  - optical lens systems, geometry of integrated circuits, aerodynamic shape, engines
- Uncertainty analysis
  - sensitivity of uncertainty variance to each contributing cause
- Markov Chain Monte Carlo (e.g., Hamiltonian method)
  - generation of random samples from a prob. dens. function

Quantification of uncertainties in simulation predictions

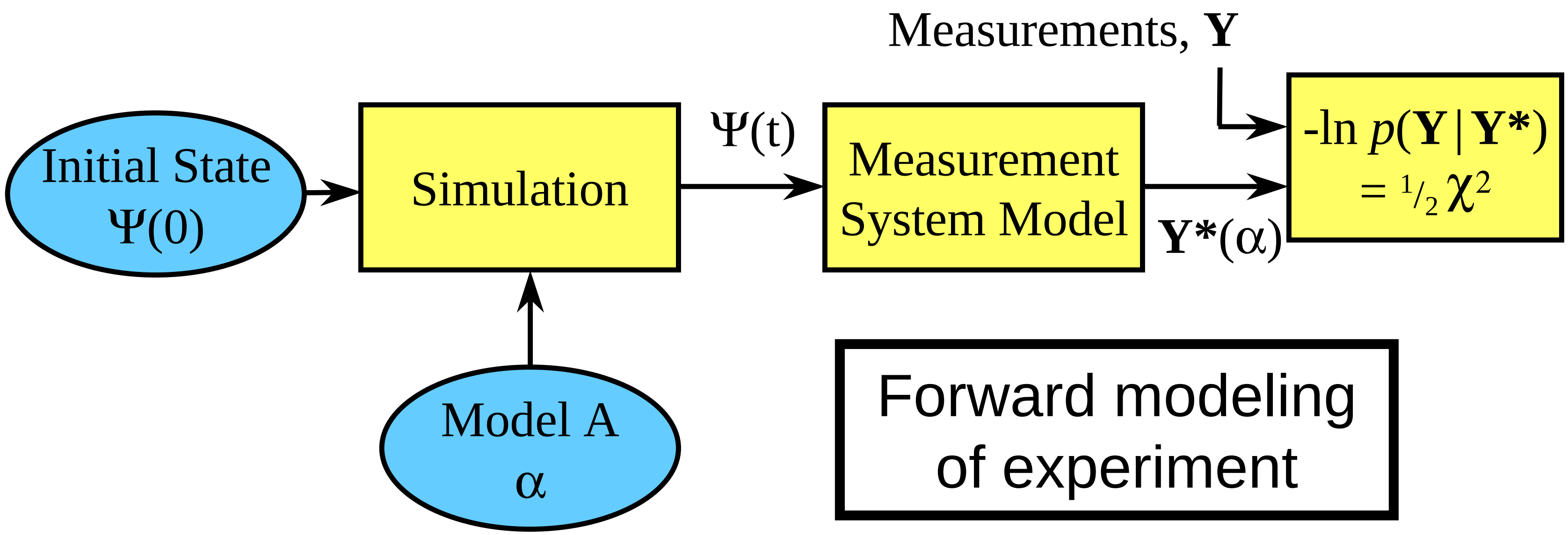
- Bayesian approach to analyzing single experiments
  - estimation of model parameters and their uncertainties
- Estimating uncertainties in simulation code predictions for new situation
- Graphical probabilistic modeling
  - analysis of numerous experiments in terms of many physical models
  - complete uncertainty analysis
  - check consistency among experiments (model checking)

Simulation code



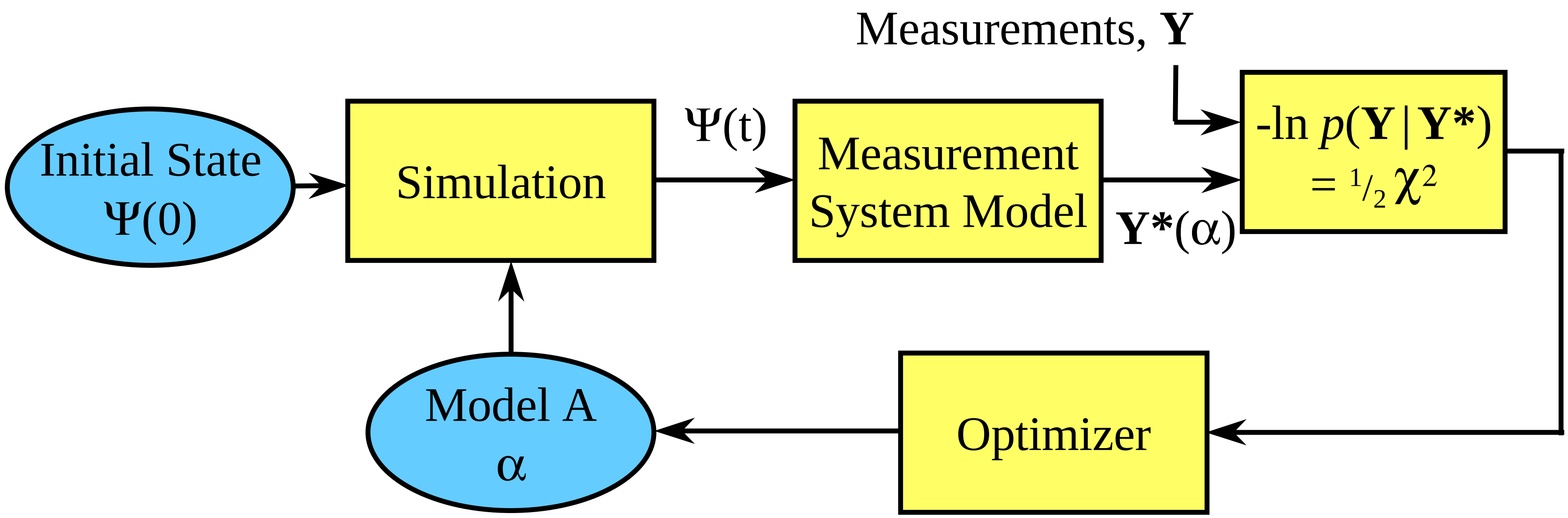
- Simulation code predicts state of time-evolving system:
  $\Psi(t)$ = time-dependent state of system
  $\Psi(0)$ = initial state of system
- Properties of one system component described by physics model A with parameter vector α (e.g., constitutive relations)

Comparison of simulation with experiment



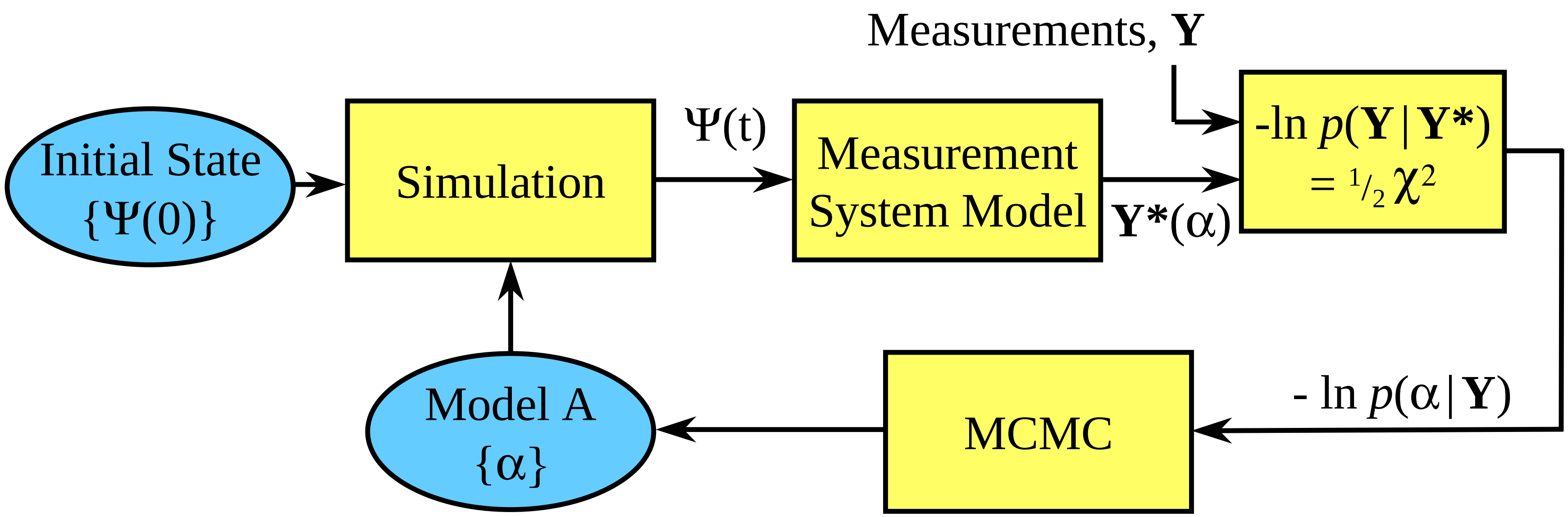
- Measurement system model transforms the simulated state of the physical system $\Psi(t)$ into measurements $\mathbf{Y^*}$ that would be obtained in the experiment
- Mismatch with data summarized by minus-log-likelihood, $-\ln p(\mathbf{Y}|\mathbf{Y^*}) = \frac{1}{2}\chi^2$

Parameter estimation - maximum likelihood



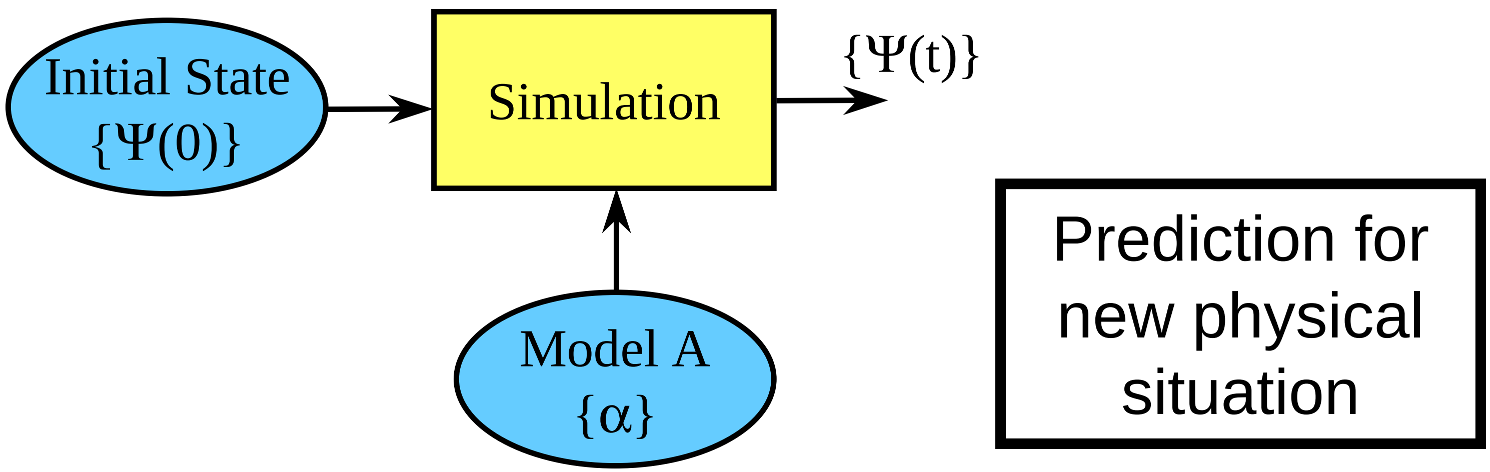
- Optimizer adjusts parameters (vector α) to minimize -ln $p(\mathbf{Y}\,|\mathbf{Y}^*(\alpha))$
- Result is maximum likelihood estimate for α (also known as minimum-chi-squared solution)
- Optimization process is accelerated by using gradient-based algorithms along with adjoint differentiation to calculate gradients of forward model

Parameter uncertainties via MCMC



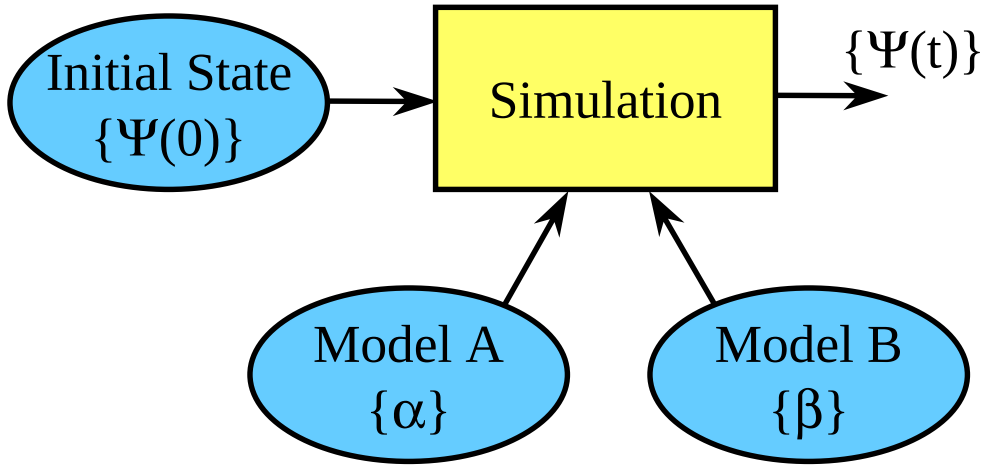
- Markov Chain Monte Carlo (MCMC) algorithm generates a random sequence of parameters that sample posterior probability of parameters for given data **Y,** $p(\alpha \mid \mathbf{Y})$, which yields plausible set of parameters $\{\alpha\}$.
- Must include uncertainty in initial state of system, $\{\Psi(0)\}$

Simulation of plausible predictions - characterize uncertainty in prediction of new situation



- Generates plausible predictions for known uncertainties in parameters
  - $\{\alpha\}$ = plausible sets of parameter vector α
  - $\{\Psi(t)\}$ = plausible sets of dynamic state of system
- Monte Carlo method - run simulation code for each random draw from pdf for α, $p(\alpha|.)$, to obtain set of predictions $\{\Psi(t)\}$

Plausible outcomes for many models

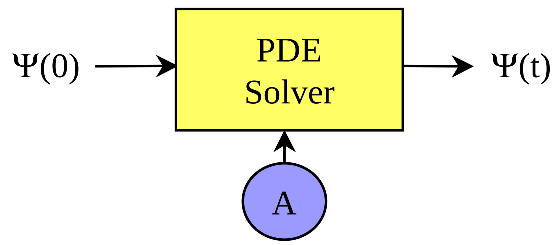


- Integrated simulation code predicts plausible results for known uncertainties in initial conditions and material models
  - $\{\alpha\}$ = plausible sets of parameter vector α for material A
  - $\{\beta\}$ = plausible sets of parameter vector β for material B
  - $\{\Psi(0)\}$ = plausible sets of initial state of system
  - $\{\Psi(t)\}$ = plausible sets of dynamic state of system

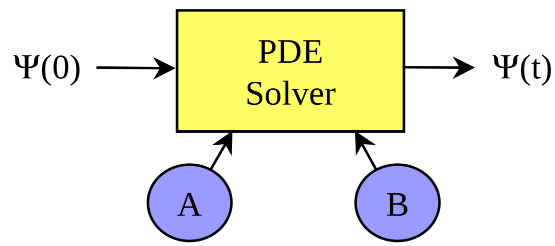
Validation Experiments

Full validation requires hierarchy of experiments

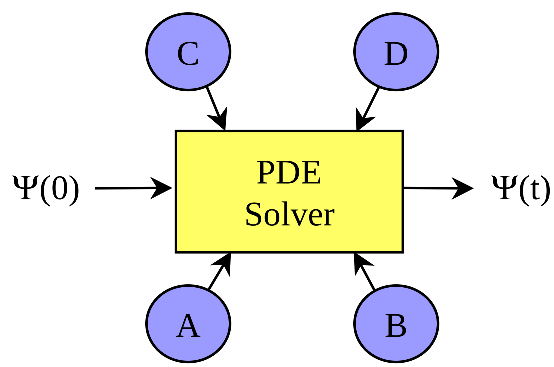
- **Basic** experiments determine individual physics models



- **Partially integrated** experiments involve combinations of two or more elemental models

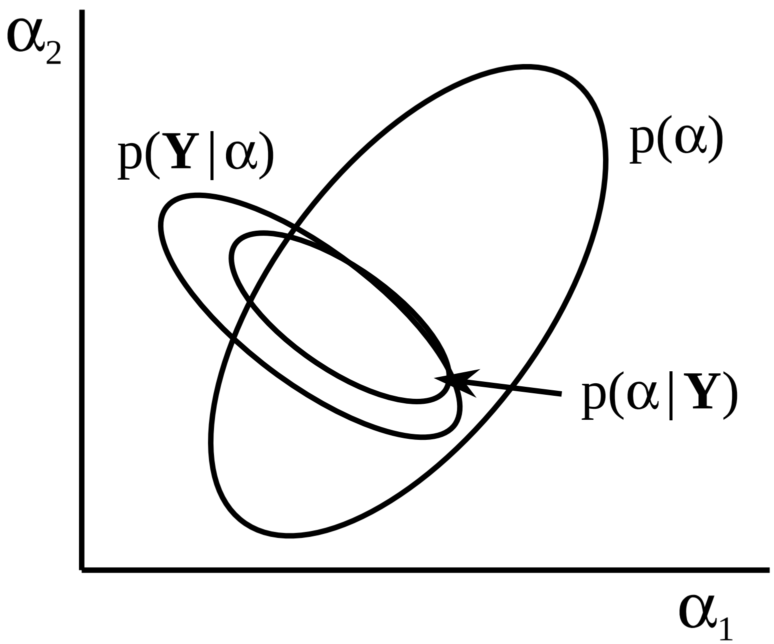


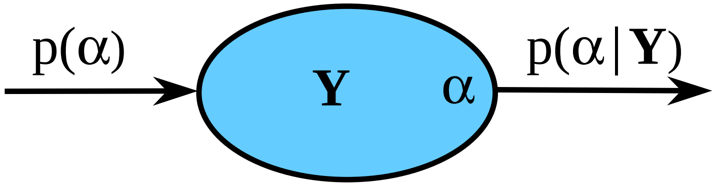
- **Fully integrated** experiments require complete set of models needed to describe final application of simulation code



Graphical probabilistic modeling

- Analysis of experimental data **Y** improves on prior knowledge about parameter vector α
- Bayes law:
  p(α | **Y**) ~ p(**Y** | α) p(α)
  (posterior ~ likelihood x prior)
- Use bubble to represent effect of analysis based on data **Y**
- In terms of logs:
  - ln p(α | **Y**) =
  - ln p(**Y** | α) - ln p(α) + constant

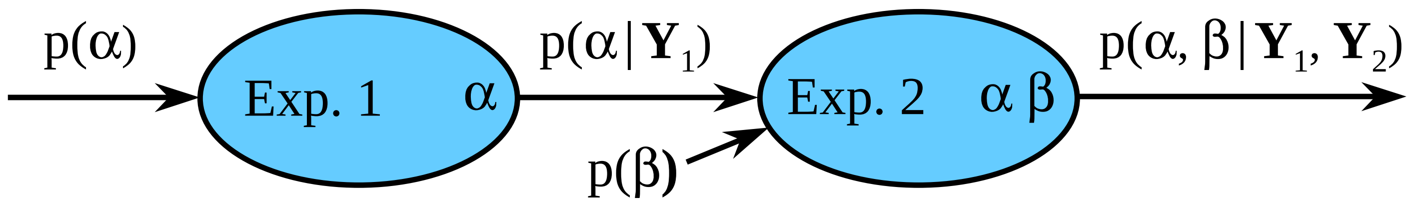




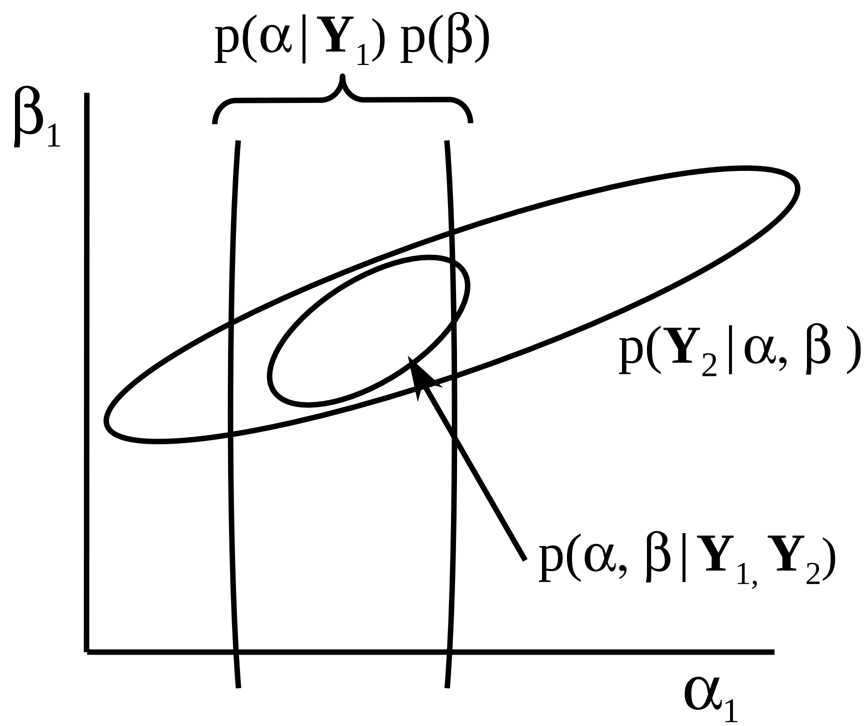
- **Not** the same as a Bayesian network

Graphical probabilistic modeling

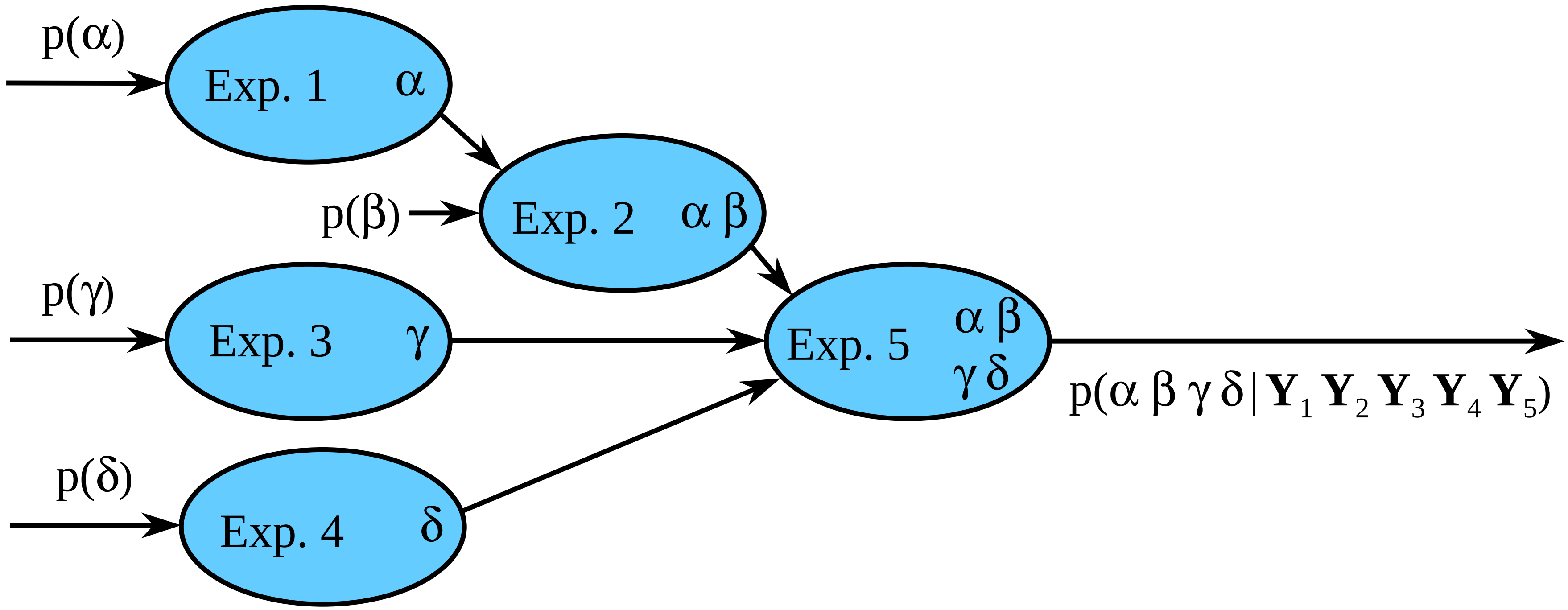
Propagate uncertainty through a sequence of analyses



- First experiment determines α, with uncertainties given by $p(\alpha|\mathbf{Y}_1)$
- Second experiment not only determines β but also refines knowledge of α
- Outcome is joint pdf in α and β, $p(\alpha, \beta|\mathbf{Y}_{1,}\mathbf{Y}_2)$ (NB: correlations)



Example of analysis of several experiments



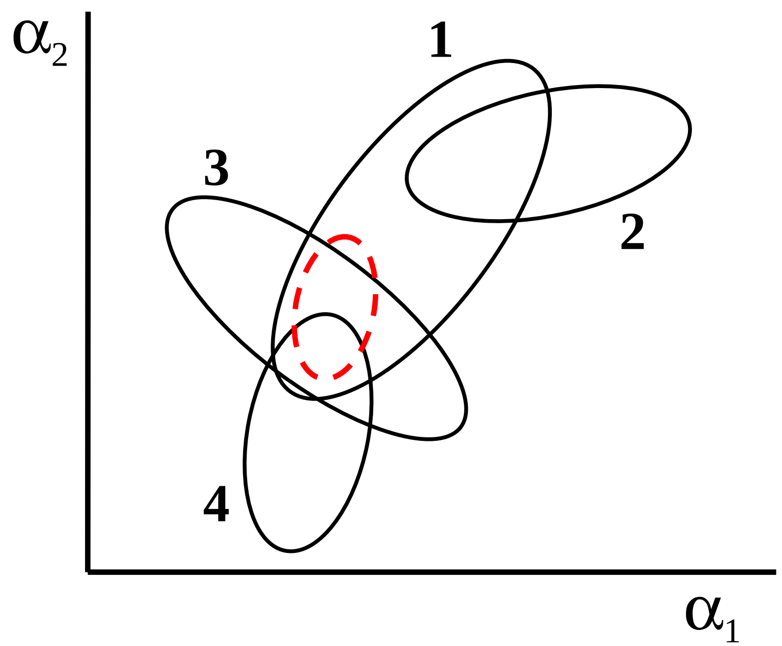
Output of final analysis is full joint probability for all parameters based on all experiments

Use of Gaussian pdfs simplifies computations

Model checking

Check that model consistent with all experimental data

- Important part of any analysis
- Check consistency of full posterior wrt. each of its contributions.
- Example shown at right:
  - likelihoods from Exps. 1 and 2 are mutually consistent
  - however, Exp. 2 is inconsistent with posterior (dashed) from all exps.
  - inconsistency must be resolved in terms of correction to model and/or interpretation of experiment



Summary

- A methodology has been presented to combine experimental results from many experiments relevant to several basic physics models in the context of a simulation code
- Propose building to implement this approach to
  - serve as a database of experiments showing links between analyses
  - permit logically consistent inferences about models based on all information
  - provide a natural way to understand limits to parameter adjustment to match data from fully integrated experiments

Summary (cont'd)

- Many challenges remain
  - systematic experimental uncertainties (effects common to many data)
  - identification and resolution of inconsistencies between experiments and simulation code
  - inclusion of other sources of uncertainty: material inhomogeneity, chaotic or turbulent behavior, numerical computation

Bibliography

- "Uncertainty assessment for reconstructions based on deformable models," K. M. Hanson et al., *Int. J. Imaging Syst. Technol.* **8**, pp. 506-512 (1997); use of MCMC
- "Operation of the Bayes Inference Engine," K. M. Hanson et al., in *Maximum Entropy and Bayesian Methods*, pp. 309-318 (Kluwer, 1999)
- "Posterior sampling with improved efficiency," K. M. Hanson et al., *Proc. SPIE* **3338**, pp. 371-382 (1998); includes introduction to MCMC
- "Inversion based on complex simulations," K. M. Hanson, *Maximum Entropy and Bayesian Methods*, pp. 121-135 (Kluwer Academic, 1998); describes adjoint differentiation and its usefulness in inverting simulations
- "A framework for assessing uncertainties in simulation predictions", K. M. Hanson, *Physica D* **133**, pp. 179-188 (2000); an integrated approach to determining uncertainties in physics modules and their effect on predictions
- "The hard truth," K. M. Hanson et al., in *Maximum Entropy and Bayesian Methods*, pp. 157-164 (Kluwer, 1996); novel technique to compute uncertainties by probing the stiffness of the posterior

These and other papers available at http://home.lanl.gov/kmh/